Determining the Epipolar Geometry and its Uncertainty: A Review

Zhengyou Zhang

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Determining the Epipolar Geometry and its Uncertainty: A Review

Zhengyou Zhang

N° 2927

July 1996

THÈME 3

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Determining the Epipolar Geometry and its Uncertainty: A Review

Zhengyou Zhang

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Robotvis

Rapport de recherche n°2927 — July 1996 — 56 pages

Abstract: Two images of a single scene/object are related by the epipolar geometry, which can be described by a 3x3 singular matrix called the essential matrix if images' internal parameters are known, or the fundamental matrix otherwise. It captures all geometric information contained in two images, and its determination is very important in many applications such as scene modeling and vehicle navigation. This paper gives an introduction to the epipolar geometry, and provides a complete review of the current techniques for estimating the fundamental matrix and its uncertainty. A well-founded measure is proposed to compare these techniques. Projective reconstruction is also reviewed. The softwares which we have developed for this review are available on the Internet.

Key-words: Epipolar Geometry, Fundamental Matrix, Calibration, Reconstruction, Parameter Estimation, Robust Techniques, Uncertainty Characterization, Performance Evaluation, Software

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Tout ce que vous voulez savoir sur le calcul de la matrice fondamentale

Résumé : La géométrie épipolaire décrit la relation entre deux images d'une même scène, qui est caractérisée par une matrice 3x3 singulière. Cette matrice est connue sous le nom de matrice essentielle si les paramètres intrinsèques des images sont connus, de matrice fondamentale s'ils ne le sont pas. Elle contient toutes les informations géométriques des deux images. Son estimation est très importante pour beaucoup d'applications comme la modélisation de scènes et le déplacement d'un robot mobile autonome. Cet article fournit une introduction à la géométrie épipolaire, et fait une revue complète des techniques existantes d'estimation de la matrice fondamentale et de son incertitude. Nous introduisons une méthode de mesure permettant de comparer avec précision ces techniques. La reconstruction projective est aussi passée en revue. Les logiciels que nous avons développés pour effectuer cette revue sont disponible sur l'Internet.

Mots-clé : Géométrie épipolaire, matrice fondamentale, calibration, reconstruction, estimation de paramètres, techniques robustes, caractérisation d'incertitude, évaluation de performance, logiciel

Contents

List of Figures

List of Tables

1 Introduction

Two perspective images of a single rigid object/scene are related by the so-called *epipolar geometry*, which can be described by a 3×3 singular matrix. If the internal (intrinsic) parameters of the images (e.g. the focal length, the coordinates of the principal point, etc) are known, we can work with the *normalized image coordinates* (Faugeras 1993), and the matrix is known as the *essential matrix* (Longuet-Higgins 1981); otherwise, we have to work with the *pixel image coordinates*, and the matrix is known as the *fundamental matrix* (Luong 1992, Faugeras 1995, Luong and Faugeras 1996). It contains all geometric information that is necessary for establishing correspondences between two images, from which three-dimensional structure of the perceived scene can be inferred. In a stereovision system where the camera geometry is calibrated, it is possible to calculate such a matrix from the camera perspective projection matrices through calibration (Ayache 1991, Faugeras 1993). When the intrinsic parameters are known but the extrinsic ones (the rotation and translation between the two images) are not, the problem is known as motion and structure from motion, and has been extensively studied in Computer Vision; two excellent reviews are already available in this domain (Aggarwal and Nandhakumar 1988, Huang and Netravali 1994). We are interested here in different techniques for estimating the fundamental matrix from two uncalibrated images, i.e. the case where both the intrinsic and extrinsic parameters of the images are unknown. From this matrix, we can reconstruct a projective structure of the scene, defined up to a 4×4 matrix transformation.

The study of uncalibrated images has many important applications. The reader may wonder the usefulness of such a projective structure. We cannot obtain any metric information from a projective structure: measurements of lengths and angles do not make sense. However, a projective structure still contains rich information, such as coplanarity, collinearity, and cross ratios (ratio of ratios of distances), which is sometimes sufficient for artificial systems, such as robots, to perform tasks such as navigation and object recognition (Shashua 1994, Zeller and Faugeras 1994, Beardsley, Zisserman and Murray 1994).

In many applications such as the reconstruction of the environment from a sequence of video images where the parameters of the video lens is submitted to continuous modification, camera calibration in the classical sense is not possible. We cannot exact any metric information, but a projective structure is still possible if the camera can be considered as a pinhole. Furthermore, if we can introduce some knowledge of the scene into the projective structure, we can obtain more specific structure of the scene. For example, by specifying a plane at infinity (in practice, we need only to specify a plane sufficiently far away), an affine structure can be computed, which preserves parallelism and ratios of distances (Quan 1993, Faugeras 1995). Hartley, Gupta and Chang (1992) first reconstruct a projective structure, and then use 8 ground reference points to obtain the Euclidean structure and the camera parameters. Mohr, Boufama and Brand (1993) embed constraints such as location of points, parallelism and vertical planes (e.g. walls) directly into a minimization procedure to determine a Euclidean structure. Robert and Faugeras (1993) show that the 3D convex hull of an object can be computed from a pair of images whose epipolar geometry is known.

If we assume that the camera parameters do not change between successive views, the projective invariants can even be used to calibrate the cameras in the classical sense without using any calibration apparatus (known as *self-calibration*) (Maybank and Faugeras 1992, Faugeras, Luong and Maybank 1992, Luong 1992, Zhang, Luong and Faugeras 1996, Enciso 1995).

Recently, we have shown (Zhang 1996a) that even in the case where images are calibrated, more reliable results can be obtained if we use the constraints arising from uncalibrated images as an intermediate step.

This paper gives an introduction to the epipolar geometry, provides a new formula of the fundamental matrix which is valid for both perspective and affine cameras, and reviews different methods reported in the literature for estimating the fundamental matrix. Furthermore, a new method is described to compare two estimations of the fundamental matrix. It is based on a measure obtained through sampling the whole visible 3D space. Projective reconstruction is also reviewed. The software called `FMatrix` which implements the reviewed methods and the software called `Fdiff` which computes the difference between two fundamental matrices are both available from my home page:

`http://www.inria.fr/robotvis/personnel/zzhang/zzhang-eng.html`

`FMatrix` detects false matches, computes the fundamental matrix and its uncertainty, and the projective reconstruction of the points as well.

2 Epipolar Geometry and Problem Statement

2.1 Notation

A camera is described by the widely used pinhole model. The coordinates of a 3D point $\mathtt{M} = [x, y, z]^T$ in a world coordinate system and its retinal image coordinates $\mathbf{m} = [u, v]^T$ are related by

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} ,$$

where s is an arbitrary scale, and $\mathbf{P}$ is a 3×4 matrix, called the perspective projection matrix. Denoting the homogeneous coordinates of a vector $\mathbf{x} = [x, y, \cdots]^T$ by $\widetilde{\mathbf{x}}$, i.e., $\widetilde{\mathbf{x}} = [x, y, \cdots, 1]^T$, we have $s\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathtt{M}}$.

The matrix $\mathbf{P}$ can be decomposed as

$$\mathbf{P} = \mathbf{A} [\mathbf{R}\ \mathbf{t}] ,$$

where $\mathbf{A}$ is a 3×3 matrix, mapping the normalized image coordinates to the retinal image coordinates, and $(\mathbf{R}, \mathbf{t})$ is the 3D displacement (rotation and translation) from the world coordinate system to the camera coordinate system.

The quantities related to the second camera is indicated by $'$. For example, if $\mathbf{m}_i$ is a point in the first image, $\mathbf{m}'_i$ denotes its corresponding point in the second image.

A line l in the image passing through point $\mathbf{m} = [u, v]^T$ is described by equation $au + bv + c = 0$. Let $\mathrm{l} = [a, b, c]^T$, then the equation can be rewritten as $\mathrm{l}^T \widetilde{\mathbf{m}} = 0$ or $\widetilde{\mathbf{m}}^T \mathrm{l} = 0$. Multiplying l by any non-zero scalar will define the same 2D line. Thus, a 2D line is represented by a homogeneous 3D vector. The distance from point $\mathbf{m}_0 = [u_0, v_0]^T$ to line $\mathrm{l} = [a, b, c]^T$ is given by

$$d(\mathbf{m}_0, \mathrm{l}) = \frac{au_0 + bv_0 + c}{\sqrt{a^2 + b^2}} .$$

Note that we here use the *signed* distance.

2.2 Epipolar Geometry and Fundamental Matrix

The epipolar geometry exists between any two camera systems. Consider the case of two cameras as shown in Fig. 1. Let C and C' be the optical centers of the first and second

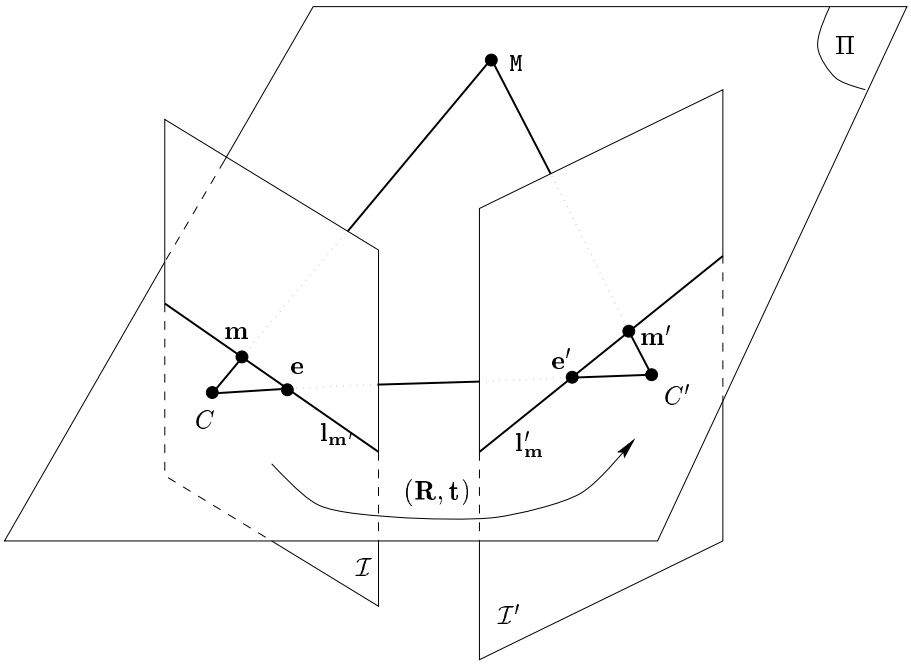

Figure 1: The epipolar geometry

cameras, respectively. Given a point $\mathbf{m}$ in the first image, its corresponding point in the second image is constrained to lie on a line called the *epipolar line* of $\mathbf{m}$, denoted by $\mathrm{l}'_{\mathbf{m}}$. The

line $\mathbf{l}'_{\mathbf{m}}$ is the intersection of the plane Π, defined by $\mathbf{m}$, C and C' (known as the *epipolar plane*), with the second image plane $\mathcal{I}'$. This is because image point $\mathbf{m}$ may correspond to an arbitrary point on the semi-line $C\mathtt{M}$ ($\mathtt{M}$ may be at infinity) and that the projection of $C\mathtt{M}$ on $\mathcal{I}'$ is the line $\mathbf{l}'_{\mathbf{m}}$. Furthermore, one observes that all epipolar lines of the points in the first image pass through a common point $\mathbf{e}'$, which is called the *epipole*. Epipole $\mathbf{e}'$ is the intersection of the line CC' with the image plane $\mathcal{I}'$. This can be easily understood as follows. For each point $\mathbf{m}_k$ in the first image $\mathcal{I}$, its epipolar line $\mathbf{l}'_{\mathbf{m}_k}$ in $\mathcal{I}'$ is the intersection of the plane Π^k, defined by $\mathbf{m}_k$, C and C', with image plane $\mathcal{I}'$. All epipolar planes Π^k thus form a pencil of planes containing the line CC'. They must intersect $\mathcal{I}'$ at a common point, which is $\mathbf{e}'$. Finally, one can easily see the symmetry of the epipolar geometry. The corresponding point in the first image of each point $\mathbf{m}'_k$ lying on $\mathbf{l}'_{\mathbf{m}_k}$ must lie on the epipolar line $\mathbf{l}_{\mathbf{m}'_k}$, which is the intersection of the same plane Π^k with the first image plane $\mathcal{I}$. All epipolar lines form a pencil containing the epipole $\mathbf{e}$, which is the intersection of the line CC' with the image plane $\mathcal{I}$. The symmetry leads to the following observation. If $\mathbf{m}$ (a point in $\mathcal{I}$) and $\mathbf{m}'$ (a point in $\mathcal{I}'$) correspond to a single physical point $\mathtt{M}$ in space, then $\mathbf{m}$, $\mathbf{m}'$, C and C' must lie in a single plane. This is the well-known *co-planarity constraint* in solving motion and structure from motion problems when the intrinsic parameters of the cameras are known (Longuet-Higgins 1981).

The computational significance in matching different views is that for a point in the first image, its correspondence in the second image must lie on the epipolar line in the second image, and then the search space for a correspondence is reduced from 2 dimensions to 1 dimension. This is called the *epipolar constraint*. Algebraically, in order for $\mathbf{m}$ in the first image and $\mathbf{m}'$ in the second image to be matched, the following equation must be satisfied:

$$\widetilde{\mathbf{m}}^T \mathbf{F} \widetilde{\mathbf{m}}' = 0 \quad \text{with } \mathbf{F} = \mathbf{A}^{-T}[\mathbf{t}]_\times \mathbf{R} \mathbf{A}'^{-1} \,, \tag{1}$$

where $(\mathbf{R}, \mathbf{t})$ is the rigid transformation (rotation and translation) which brings points expressed in the *second* camera coordinate system to the first one, and $[\mathbf{t}]_\times$ is the antisymmetric matrix defined by $\mathbf{t}$ such that $[\mathbf{t}]_\times \mathbf{x} = \mathbf{t} \times \mathbf{x}$ for all 3D vector $\mathbf{x}$. This equation can be derived as follows. Without loss of generality, we assume that the world coordinate system coincides with the *second* camera coordinate system. From the pinhole model, we have

$$s\widetilde{\mathbf{m}} = \mathbf{A}\,[\mathbf{R}\ \mathbf{t}]\,\widetilde{\mathtt{M}}' \qquad \text{and} \qquad s'\widetilde{\mathbf{m}}' = \mathbf{A}'\,[\mathbf{I}\ \mathbf{0}]\,\widetilde{\mathtt{M}}' \,.$$

Eliminating $\mathtt{M}'$, s and s' in the above two equations, we obtain equation (1). Geometrically, $\mathbf{F}\widetilde{\mathbf{m}}'$ defines the epipolar line $\mathbf{l}_{\mathbf{m}'}$ of point $\mathbf{m}'$ in the first image. Equation (1) says no more than that the correspondence in the first image of point $\mathbf{m}'$ lies on the corresponding epipolar line $\mathbf{l}_{\mathbf{m}'}$. Transposing (1) yields the symmetric relation from the first image to the second image.

The 3×3 matrix $\mathbf{F}$ is called the *fundamental matrix*. Since $\det([\mathbf{t}]_\times) = 0$,

$$\det(\mathbf{F}) = 0 \,. \tag{2}$$

$\mathbf{F}$ is of rank 2. Besides, it is only defined up to a scalar factor, because if $\mathbf{F}$ is multiplied by an arbitrary scalar, equation (1) still holds. Therefore, a fundamental matrix has only seven degrees of freedom. There are only 7 independent parameters among the 9 elements of the fundamental matrix.

Convention note: We use the second camera coordinate system as the world coordinate system, which is in accordance with the convention used in (Faugeras 1993). Several researchers prefer to use the first camera coordinate system, then (1) becomes $\widetilde{\mathbf{m}}'^T\mathbf{F}'\widetilde{\mathbf{m}} = 0$ with $\mathbf{F}' = [\mathbf{t}']_\times\mathbf{R}'$, where $(\mathbf{R}', \mathbf{t}')$ transforms points from the first camera coordinate system to the second. The relation between $(\mathbf{R}, \mathbf{t})$ and $(\mathbf{R}', \mathbf{t}')$ is given by $\mathbf{R}' = \mathbf{R}^T$, and $\mathbf{t}' = -\mathbf{R}^T\mathbf{t}$. The reader can easily verify that $\mathbf{F} = \mathbf{F}'^T$.

2.3 A General Form of Epipolar Equation for Any Projection Model

In this section we will derive a general form of epipolar equation which does not assume any particular projection model (Xu and Zhang 1996).

A point $\mathbf{m}$ in the first image is matched to a point $\mathbf{m}'$ in the second image. From the camera projection model (orthographic, weak perspective, affine, or full perspective), we have $s\widetilde{\mathbf{m}} = \mathbf{P}\widetilde{\mathtt{M}}'$ and $s'\widetilde{\mathbf{m}}' = \mathbf{P}'\widetilde{\mathtt{M}}'$, where $\mathbf{P}$ and $\mathbf{P}'$ are 3×4 matrices. An image point $\mathbf{m}'$ defines actually an optical ray, on which every space point $\widetilde{\mathtt{M}}'$ projects on the second image at $\widetilde{\mathbf{m}}'$. This optical ray can be written in parametric form as

$$\widetilde{\mathtt{M}}' = s'\mathbf{P}'^{+}\widetilde{\mathbf{m}}' + \mathbf{p}'^{\perp} , \tag{3}$$

where $\mathbf{P}'^{+}$ is the pseudo-inverse of matrix $\mathbf{P}'$:

$$\mathbf{P}'^{+} = \mathbf{P}'^{T}(\mathbf{P}'\mathbf{P}'^{T})^{-1} , \tag{4}$$

and $\mathbf{p}'^{\perp}$ is any 4-vector that is perpendicular to all the *row* vectors of $\mathbf{P}'$, i.e.

$$\mathbf{P}'\mathbf{p}'^{\perp} = \mathbf{0} .$$

Thus, $\mathbf{p}'^{\perp}$ is a null vector of $\mathbf{P}'$. As a matter of fact, $\mathbf{p}'^{\perp}$ indicates the position of the optical center (to which all optical rays converge). We show later how to determine $\mathbf{p}'^{\perp}$. For a particular value s', equation (3) corresponds to a point on the optical ray defined by $\mathbf{m}'$. Equation (3) is easily justified by projecting $\mathtt{M}'$ onto the second image, which indeed gives $\mathbf{m}'$.

Similarly, an image point $\mathbf{m}$ in the first image defines also an optical ray. Requiring the two rays to intersect in space implies that a point $\mathtt{M}'$ corresponding to a particular s' in (3) must project onto the first image at $\mathbf{m}$, that is

$$s\widetilde{\mathbf{m}} = s'\mathbf{P}\mathbf{P}'^{+}\widetilde{\mathbf{m}}' + \mathbf{P}\mathbf{p}'^{\perp} .$$

Performing a cross product with $\mathbf{P}\mathbf{p}'^{\perp}$ yields

$$s(\mathbf{P}\mathbf{p}'^{\perp}) \times \widetilde{\mathbf{m}} = s'(\mathbf{P}\mathbf{p}'^{\perp}) \times (\mathbf{P}\mathbf{P}'^{+}\widetilde{\mathbf{m}}') \ .$$

Eliminating s and s' by multiplying $\widetilde{\mathbf{m}}^T$ from the left (equivalent to a dot product), we have

$$\widetilde{\mathbf{m}}^T \mathbf{F} \widetilde{\mathbf{m}}' = 0 \ , \tag{5}$$

where $\mathbf{F}$ is a 3×3 matrix, called *fundamental matrix*:

$$\mathbf{F} = [\mathbf{P}\mathbf{p}'^{\perp}]_{\times} \mathbf{P}\mathbf{P}'^{+} \ . \tag{6}$$

It can also be shown that this expression is equivalent to (1) for the full perspective projection (see Xu and Zhang 1996), but it is more general. Indeed, (1) assumes that the first 3×3 sub-matrix of $\mathbf{P}'$ is invertible, and thus is only valid for full perspective projection but not for affine cameras (see Sect. 5.3), while (6) makes use of the pseudoinverse of the projection matrix, which is valid for both full perspective projection as well as affine cameras. Therefore the equation does not depend on any specific knowledge of projection model. Replacing the projection matrix in the equation by specific projection matrix for each specific projection model (e.g. orthographic, weak perspective, affine or full perspective) produces the epipolar equation for that specific projection model. See (Xu and Zhang 1996) for more details.

The vector $\mathbf{p}'^{\perp}$ still needs to be determined. We first note that such a vector must exist because the difference between the row dimension and the column dimension is one, and that the row vectors are generally independent from each other. Indeed, one way to obtain $\mathbf{p}'^{\perp}$ is

$$\mathbf{p}'^{\perp} = (\mathbf{I} - \mathbf{P}'^{+}\mathbf{P}')\boldsymbol{\omega} \ , \tag{7}$$

where $\boldsymbol{\omega}$ is an arbitrary 4-vector. To show that $\mathbf{p}'^{\perp}$ is perpendicular to each row of $\mathbf{P}'$, we multiply $\mathbf{p}'^{\perp}$ by $\mathbf{P}'$ from the left: $\mathbf{P}'\mathbf{p}'^{\perp} = (\mathbf{P}' - \mathbf{P}'\mathbf{P}'^{T}(\mathbf{P}'\mathbf{P}'^{T})^{-1}\mathbf{P}')\boldsymbol{\omega} = \mathbf{0}$, which is indeed a zero vector. The action of $\mathbf{I} - \mathbf{P}'^{+}\mathbf{P}'$ is to transform an arbitrary vector to a vector that is perpendicular to every row vector of $\mathbf{P}'$. If $\mathbf{P}'$ is of rank 3 (which is the case for both perspective and affine cameras), then $\mathbf{p}'^{\perp}$ is unique up to a scale factor.

2.4 Problem Statement

The problem considered in the sequel is the estimation of $\mathbf{F}$ from a sufficiently large set of point correspondences: $\{(\mathbf{m}_i, \mathbf{m}'_i) \,|\, i = 1, \ldots, n\}$, where $n \geq 7$. The point correspondences between two images can be established by a technique such as that described in (Zhang, Deriche, Faugeras and Luong 1995). We allow, however, that a fraction of the matches may be incorrectly paired, and thus the estimation techniques should be robust.

3 Techniques for Estimating the Fundamental Matrix

Let a point $\mathbf{m}_i = [u_i, v_i]^T$ in the first image be matched to a point $\mathbf{m}'_i = [u'_i, v'_i]^T$ in the second image. They must satisfy the epipolar equation (1), i.e. $\widetilde{\mathbf{m}}_i^T \mathbf{F} \widetilde{\mathbf{m}}'_i = 0$. This equation can be written as a linear and homogeneous equation in the 9 unknown coefficients of matrix $\mathbf{F}$:

$$\mathbf{u}_i^T \mathbf{f} = 0 \ , \tag{8}$$

where

$$\begin{aligned} \mathbf{u}_i &= [u_i u'_i, u_i v'_i, u_i, v_i u'_i, v_i v'_i, v_i, u'_i, v'_i, 1]^T \\ \mathbf{f} &= [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \ , \end{aligned}$$

where F_{ij} is the element of $\mathbf{F}$ at row i and column j.

If we are given n point matches, by stacking (8), we have the following linear system to solve:

$$\mathbf{U}_n \mathbf{f} = \mathbf{0} \ ,$$

where

$$\mathbf{U}_n = [\mathbf{u}_1, \cdots, \mathbf{u}_n]^T \ .$$

This set of linear homogeneous equations, together with the rank constraint of the matrix $\mathbf{F}$, allow us to estimate the epipolar geometry.

3.1 Exact Solution with 7 Point Matches

As described in Sect. 2.2, a fundamental matrix $\mathbf{F}$ has only 7 degrees of freedom. Thus, 7 is the minimum number of point matches required for having a solution of the epipolar geometry.

In this case, $n = 7$ and $\text{rank}(\mathbf{U}_7) = 7$. Through singular value decomposition, we obtain vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ which span the null space of $\mathbf{U}_7$. The null space is a linear combination of $\mathbf{f}_1$ and $\mathbf{f}_2$, which correspond to matrices $\mathbf{F}_1$ and $\mathbf{F}_2$, respectively. Because of its homogeneity, the fundamental matrix is a one-parameter family of matrices $\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2$. Since the determinant of $\mathbf{F}$ must be null, i.e.

$$\det[\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2] = 0 \ ,$$

we obtain a cubic polynomial in α. The maximum number of real solutions is 3. For each solution α, the fundamental matrix is then given by

$$\mathbf{F} = \alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2 \ .$$

Actually, this technique has already been used in estimating the essential matrix when 7 point matches in normalized coordinates are available (Huang and Netravali 1994). It is also mentioned in (Torr 1995) for estimating the fundamental matrix.

As a matter of fact, the result that there may have three solutions given 7 matches has been known since 1800's (Hesse 1863, Sturm 1869). Sturm's algorithm (Sturm 1869) computes the epipoles and the epipolar transformation (see Sect. 2.2) from 7 point matches. It is based on the observation that the epipolar lines in the two images are related by a homography, and thus the cross-ratios of four epipolar lines is invariant. In each image, the 7 points define 7 lines going through the unknown epipole, thus providing 4 independent cross-ratios. Since these cross-ratios should remain the same in the two images, one obtains 4 cubic polynomial equations in the coordinates of the epipoles (4 independent parameters). It is shown that there may exist up to three solutions for the epipoles.

3.2 Analytic Method with 8 or More Point Matches

In practice, we are given more than 7 matches. If we ignore the rank-2 constraint, we can use a least-squares method to solve

$$\min_{\mathbf{F}} \sum_i (\widetilde{\mathbf{m}}_i^T \mathbf{F} \widetilde{\mathbf{m}}_i')^2 \ , \tag{9}$$

which can be rewritten as:

$$\min_{\mathbf{f}} \|\mathbf{U}_n \mathbf{f}\|^2 \ . \tag{10}$$

The vector $\mathbf{f}$ is only defined up to an unknown scale factor. The trivial solution $\mathbf{f}$ to the above problem is $\mathbf{f} = \mathbf{0}$, which is not what we want. To avoid it, we need to impose some constraint on the coefficients of the fundamental matrix. Several methods are possible and are presented below. We will call them *the 8-point algorithm*, although more than 8 point matches can be used.

3.2.1 Linear Least-Squares Technique

The first method sets one of the coefficients of $\mathbf{F}$ to 1, and then solves the above problem using linear least-squares techniques. Without loss of generality, we assume that the last element of vector $\mathbf{f}$ (i.e. $f_9 = F_{33}$) is not equal to zero, and thus we can set $f_9 = -1$. This gives

$$\|\mathbf{U}_n \mathbf{f}\|^2 = \|\mathbf{U}'_n \mathbf{f}' - \mathbf{c}_9\|^2 = \mathbf{f}'^T \mathbf{U}_n'^T \mathbf{U}'_n \mathbf{f}' - 2\mathbf{c}_9^T \mathbf{U}'_n \mathbf{f}' + \mathbf{c}_9^T \mathbf{c}_9 \ ,$$

where $\mathbf{U}'_n$ is the $n \times 8$ matrix composed of the first 8 columns of $\mathbf{U}_n$, and $\mathbf{c}_9$ is the ninth column of $\mathbf{U}_n$. The solution is obtained by requiring the first derivative to be zero, i.e.

$$\frac{\partial \|\mathbf{U}_n \mathbf{f}\|^2}{\partial \mathbf{f}'} = \mathbf{0} \ .$$

By definition of vector derivatives, $\partial(\mathbf{a}^T\mathbf{x})/\partial\mathbf{x} = \mathbf{a}$, for all vector $\mathbf{a}$. We thus have

$$2\mathbf{U}_n'^T\mathbf{U}'_n\mathbf{f}' - 2\mathbf{U}_n'^T\mathbf{c}_9 = \mathbf{0} \ , \quad \text{or} \quad \mathbf{f}' = \left(\mathbf{U}_n'^T\mathbf{U}'_n\right)^{-1}\mathbf{U}_n'^T\mathbf{c}_9 \ .$$

The problem with this method is that we do not know a priori which coefficient is not zero. If we set an element to 1 which is actually zero or much smaller than the other elements, the result will be catastrophic. A remedy is to try all nine possibilities by setting one of the nine coefficients of $\mathbf{F}$ to 1 and retain the best estimation.

3.2.2 Eigen Analysis

The second method consists in imposing a constraint on the norm of $\mathbf{f}$, and in particular we can set $\|\mathbf{f}\| = 1$. Compared to the previous method, no coefficient of $\mathbf{F}$ prevails over the others. In this case, the problem (10) becomes a classical one:

$$\min_{\mathbf{f}} \|\mathbf{U}_n\mathbf{f}\|^2 \quad \text{subject to } \|\mathbf{f}\| = 1 \ . \tag{11}$$

It can be transformed into an unconstrained minimization problem through Lagrange multipliers:

$$\min_{\mathbf{f}} \mathcal{F}(\mathbf{f}, \lambda) \ , \tag{12}$$

where

$$\mathcal{F}(\mathbf{f}, \lambda) = \|\mathbf{U}_n\mathbf{f}\|^2 + \lambda(1 - \|\mathbf{f}\|^2) \tag{13}$$

and λ is the Lagrange multiplier. By requiring the first derivative of $\mathcal{F}(\mathbf{f}, \lambda)$ with respect to $\mathbf{f}$ to be zero, we have

$$\mathbf{U}_n^T\mathbf{U}_n\mathbf{f} = \lambda\mathbf{f} \ .$$

Thus, the solution $\mathbf{f}$ must be a unit eigenvector of the 9×9 matrix $\mathbf{U}_n^T\mathbf{U}_n$ and λ is the corresponding eigenvalue. Since matrix $\mathbf{U}_n^T\mathbf{U}_n$ is symmetric and positive semi-definite, all its eigenvalues are real and positive or zero. Without loss of generality, we assume the nine eigenvalues of $\mathbf{U}_n^T\mathbf{U}_n$ are in non-increasing order:

$$\lambda_1 \geq \cdots \geq \lambda_i \geq \cdots \geq \lambda_9 \geq 0 \ .$$

We therefore have 9 potential solutions: $\lambda = \lambda_i$ for $i = 1, \ldots, 9$. Back substituting the solution to (13) gives

$$\mathcal{F}(\mathbf{f}, \lambda_i) = \lambda_i \ .$$

Since we are seeking to minimize $\mathcal{F}(\mathbf{f}, \lambda)$, the solution to (11) is evidently the unit eigenvector of matrix $\mathbf{U}_n^T \mathbf{U}_n$ associated to the *smallest* eigenvalue, i.e. λ_9.

3.2.3 Imposing the Rank-2 Constraint

The advantage of the linear criterion is that it yields an analytic solution. However, we have found that it is quite sensitive to noise, even with a large set of data points. One reason is that the rank-2 constraint (i.e. $\det \mathbf{F} = 0$) is not satisfied. We can impose this constraint a posteriori. The most convenient way is to replace the matrix $\mathbf{F}$ estimated with any of the above methods by the matrix $\hat{\mathbf{F}}$ which minimizes the Frobenius norm (see Sect. B) of $\mathbf{F} - \hat{\mathbf{F}}$ subject to the constraint $\det \hat{\mathbf{F}} = 0$. Let

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

be the singular value decomposition of matrix $\mathbf{F}$, where $\mathbf{S} = \mathrm{diag}\,(\sigma_1, \sigma_2, \sigma_3)$ is a diagonal matrix satisfying $\sigma_1 \geq \sigma_2 \geq \sigma_3$ (σ_i is the i^{th} singular value), and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. It can be shown that

$$\hat{\mathbf{F}} = \mathbf{U}\hat{\mathbf{S}}\mathbf{V}^T$$

with $\hat{\mathbf{S}} = \mathrm{diag}\,(\sigma_1, \sigma_2, 0)$ minimizes the Frobenius norm of $\mathbf{F} - \hat{\mathbf{F}}$ (see the appendix Sect. B for the proof). (This method was used by Tsai and Huang (1984) in estimating the essential matrix, and is introduced by Hartley (1995) to estimate the fundamental matrix.)

3.2.4 Geometric Interpretation of the Linear Criterion

Another problem with the linear criterion is that the quantity we are minimizing is not physically meaningful. A physically meaningful quantity should be something measured in the image plane, because the available information (2D points) are extracted from images. One such quantity is the distance from a point $\mathbf{m}_i$ to its corresponding epipolar line $\mathrm{l}_i = \mathbf{F}\tilde{\mathbf{m}}'_i \equiv [l_1, l_2, l_3]^T$, which is given by (see Sect. 2.1)

$$d(\mathbf{m}_i, \mathrm{l}_i) = \frac{\tilde{\mathbf{m}}_i^T \mathrm{l}_i}{\sqrt{l_1^2 + l_2^2}} = \frac{1}{c_i} \tilde{\mathbf{m}}_i^T \mathbf{F} \tilde{\mathbf{m}}'_i \ , \tag{14}$$

where $c_i = \sqrt{l_1^2 + l_2^2}$. Thus, the criterion (9) can be rewritten as

$$\min_{\mathbf{F}} \sum_{i=1}^{n} c_i^2 d^2(\mathbf{m}_i, \mathrm{l}_i) \ .$$

This means that we are minimizing not only a physical quantity $d(\mathbf{m}_i, \mathbf{l}_i)$, but also c_i which is not physically meaningful. Luong (1992) shows that the linear criterion introduces a bias and tends to bring the epipoles towards the image center.

3.2.5 Normalizing Input Data

Hartley (1995) has analyzed, from numerical computation point of view, the high instability of this linear method if pixel coordinates are directly used, and proposed to perform a simple normalization of input data prior to running the 8-point algorithm. This technique indeed produces much better results, and is summarized below.

Suppose that coordinates $\mathbf{m}_i$ in one image are replaced by $\hat{\mathbf{m}}_i = \mathbf{T}\tilde{\mathbf{m}}_i$, and coordinates $\mathbf{m}'_i$ in the other image are replaced by $\hat{\mathbf{m}}'_i = \mathbf{T}'\tilde{\mathbf{m}}'_i$, where $\mathbf{T}$ and $\mathbf{T}'$ are any 3×3 matrices. Substituting in the equation $\tilde{\mathbf{m}}_i^T \mathbf{F} \tilde{\mathbf{m}}'_i = 0$, we derive the equation $\hat{\mathbf{m}}_i^T \mathbf{T}^{-T} \mathbf{F} \mathbf{T}'^{-1} \hat{\mathbf{m}}'_i = 0$. This relation implies that $\mathbf{T}^{-T}\mathbf{F}\mathbf{T}'^{-1}$ is the fundamental matrix corresponding to the point correspondences $\hat{\mathbf{m}}_i \leftrightarrow \hat{\mathbf{m}}'_i$. Thus, an alternative method of finding the fundamental matrix is as follows:

1. Transform the image coordinates according to transformations $\hat{\mathbf{m}}_i = \mathbf{T}\tilde{\mathbf{m}}_i$ and $\hat{\mathbf{m}}'_i = \mathbf{T}'\tilde{\mathbf{m}}'_i$.
2. Find the fundamental matrix $\hat{\mathbf{F}}$ corresponding to the matches $\hat{\mathbf{m}}_i \leftrightarrow \hat{\mathbf{m}}'_i$.
3. Retrieve the original fundamental matrix as $\mathbf{F} = \mathbf{T}^T \hat{\mathbf{F}} \mathbf{T}'$.

The question now is how to choose the transformations $\mathbf{T}$ and $\mathbf{T}'$.

Consider the second method described above, which consists in finding the eigenvector of the 9×9 matrix $\mathbf{U}_n^T \mathbf{U}_n$ associated with the least eigenvalue (for simplicity, this vector is called the *least eigenvector* in the sequel). This matrix can be expressed as $\mathbf{U}_n^T \mathbf{U}_n = \mathbf{U}\mathbf{D}\mathbf{U}^T$, where $\mathbf{U}$ is orthogonal and $\mathbf{D}$ is diagonal whose diagonal entries λ_i $(i = 1, \ldots, 9)$ are assumed to be in non-increasing order. In this case, the least eigenvector of $\mathbf{U}_n^T \mathbf{U}_n$ is the last column of $\mathbf{U}$. Denote by κ the ratio λ_1/λ_8. The parameter κ is the *condition number*[1] of the matrix $\mathbf{U}_n^T \mathbf{U}_n$, well known to be an important factor in the analysis of stability of linear problems (Golub and van Loan 1989). If κ is large, then very small changes to the data can cause large changes to the solution. The sensitivity of invariant subspaces is discussed in detail in (Golub and van Loan 1989, p.413).

The major reason for the poor condition of the matrix $\mathbf{U}_n^T \mathbf{U}_n \equiv \mathbf{X}$ is the lack of homogeneity in the image coordinates. In an image of dimension 200×200, a typical image point will be of the form $(100, 100, 1)$. If both $\tilde{\mathbf{m}}_i$ and $\tilde{\mathbf{m}}'_i$ are of this form, then $\mathbf{u}_i$ will be of the form $[10^4, 10^4, 10^2, 10^4, 10^4, 10^2, 10^2, 10^2, 1]^T$. The contribution to the matrix $\mathbf{X}$ is of the form $\mathbf{u}_i \mathbf{u}_i^T$, which will contain entries ranging between 10^8 and 1. The diagonal entries of $\mathbf{X}$ will be of the form $[10^8, 10^8, 10^4, 10^8, 10^8, 10^4, 10^4, 10^4, 1]^T$. Summing over all point matches will result in a matrix $\mathbf{X}$ whose diagonal entries are approximately in this

[1] Strictly speaking, λ_1/λ_9 is the condition number, but λ_1/λ_8 is the parameter of importance here.

proportion. We denote by $\mathbf{X}_r$ the trailing $r \times r$ principal submatrix (that is the last r columns and rows) of $\mathbf{X}$, and by $\lambda_i(\mathbf{X}_r)$ its i^{th} largest eigenvalue. Thus $\mathbf{X}_9 = \mathbf{X} = \mathbf{U}_n^T\mathbf{U}_n$ and $\kappa = \lambda_1(\mathbf{X}_9)/\lambda_8(\mathbf{X}_9)$. First, we consider the eigenvalues of $\mathbf{X}_2$. Since the sum of the two eigenvalues is equal to the trace, we see that $\lambda_1(\mathbf{X}_2) + \lambda_2(\mathbf{X}_2) = \text{trace}(\mathbf{X}_2) = 10^4 + 1$. Since eigenvalues are non-negative, we know that $\lambda_1(\mathbf{X}_2) \leq 10^4 + 1$. From the *interlacing property* (Golub and van Loan 1989, p.411), we arrive that

$$\lambda_8(\mathbf{X}_9) \leq \lambda_7(\mathbf{X}_8) \leq \cdots \leq \lambda_1(\mathbf{X}_2) \leq 10^4 + 1 \ .$$

On the other hand, also from the interlacing property, we know that the largest eigenvalue of $\mathbf{X}$ is not less than the largest diagonal entry, i.e. $\lambda_1(\mathbf{X}_9) \geq 10^8$. Therefore, the ratio $\kappa = \lambda_1(\mathbf{X}_9)/\lambda_8(\mathbf{X}_9) \geq 10^8/(10^4 + 1)$. In fact, $\lambda_8(\mathbf{X}_9)$ will usually be much smaller than $10^4 + 1$ and the condition number will be far greater. This analysis shows that *scaling the coordinates so that they are on the average equal to* unity *will improve the condition of the matrix* $\mathbf{U}_n^T\mathbf{U}_n$.

Now consider the effect of translation. A usual practice is to fix the origin of the image coordinates at the top left hand corner of the image, so that all the image coordinates are positive. In this case, *an improvement in the condition of the matrix may be achieved by translating the points so that the centroid of the points is at the origin.* Informally, if the first image coordinates (the u-coordinates) of a set of points are $\{1001.5, 1002.3, 998.7, \ldots\}$, then the significant values of the coordinates are obscured by the coordinate offset of 1000. By translating by 1000, these numbers are changed to $\{1.5, 2.3, -1.3, \ldots\}$. The significant values become now prominent.

Based on the above analysis, Hartley (1995) propose an isotropic scaling of the input data:

1. As a first step, the points are translated so that their centroid is at the origin.
2. Then, the coordinates are scaled, so that on the average a point $\widetilde{\mathbf{m}}_i$ is of the form $\widetilde{\mathbf{m}}_i = [1, 1, 1]^T$. Such a point will lie at a distance $\sqrt{2}$ from the origin. Rather than choosing different scale factors for u and v coordinates, we choose to scale the points isotropically so that the average distance from the origin to these points is equal to $\sqrt{2}$.

Such a transformation is applied to each of the two images independently.

An alternative to the isotropic scaling is an affine transformation so that the two principal moments of the set of points are both equal to unity. However, Hartley (1995) found that the results obtained were little different from those obtained using the isotropic scaling method.

Beardsley et al. (1994) mention a normalization scheme which assumes some knowledge of camera parameters. Actually, if approximate intrinsic parameters (i.e. the intrinsic matrix $\mathbf{A}$) of a camera are available, we can apply the transformation $\mathbf{T} = \mathbf{A}^{-1}$ to obtain a "quasi-Euclidean" frame.

Boufama and Mohr (1995) use implicitly data normalization by selecting 4 points, which are largely spread in the image (i.e. most distant from each other), to form a projective basis.

3.3 Analytic Method with Rank-2 Constraint

The method described in this section is due to Faugeras (1995) which imposes the rank-2 constraint during the minimization but still yields an analytic solution. Without loss of generality, let $\mathbf{f} = [\mathbf{g}^T, f_8, f_9]^T$, where $\mathbf{g}$ is a vector containing the first seven components of $\mathbf{f}$. Let $\mathbf{c}_8$ and $\mathbf{c}_9$ be the last two column vectors of $\mathbf{U}_n$, and $\mathbf{B}$ be the $n \times 7$ matrix composed of the first seven columns of $\mathbf{U}_n$. From $\mathbf{U}_n\mathbf{f} = \mathbf{0}$, we have

$$\mathbf{B}\mathbf{g} = -f_8\mathbf{c}_8 - \mathbf{f}_9\mathbf{c}_9 \ .$$

Assume that the rank of $\mathbf{B}$ is 7, we can solve for $\mathbf{g}$ by least-squares as

$$\mathbf{g} = -f_8(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{c}_8 - f_9(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{c}_9 \ .$$

The solution depends on two free parameters f_8 and f_9. As in Sect. 3.1, we can use the constraint $\det(\mathbf{F}) = 0$, which gives a third-degree homogeneous equation in f_8 and f_9, and we can solve for their ratio. Because a third-degree equation has at least one real root, we are guaranteed to obtain at least one solution for $\mathbf{F}$. This solution is defined up to a scale factor, and we can normalize $\mathbf{f}$ such that its vector norm is equal to 1. If there are three real roots, we choose the one that minimizes the vector norm of $\mathbf{U}_n\mathbf{f}$ subject to $\|\mathbf{f}\| = 1$. In fact, we can do the same computation for any of the 36 choices of pairs of coordinates of $\mathbf{f}$ and choose, among the possibly 108 solutions, the one that minimizes the previous vector norm.

The difference between this method and those described in Sect. 3.2 is that the latter impose the rank-2 constraint after application of the linear least-squares. We have experimented this method with a limited number of data sets, and found the results comparable with those obtained by the previous one.

3.4 Nonlinear Method Minimizing Distances of Points to Epipolar Lines

As discussed in Sect. 3.2.4, the linear method (10) does not minimize a physically meaningful quantity. A natural idea is then to minimize the distances between points and their corresponding epipolar lines: $\min_{\mathbf{F}} \sum_i d^2(\tilde{\mathbf{m}}_i, \mathbf{F}\tilde{\mathbf{m}}'_i)$, where $d(\cdot,\cdot)$ is given by (14). However, unlike the case of the linear criterion, the two images do not play a symmetric role. This is because the above criterion determines only the epipolar lines in the first image. As we have seen in Sect. 2.2, by exchanging the role of the two images, the fundamental matrix is changed to its transpose. To avoid the inconsistency of the epipolar geometry between the two images, we minimize the following criterion

$$\min_{\mathbf{F}} \sum_i \left(d^2(\tilde{\mathbf{m}}_i, \mathbf{F}\tilde{\mathbf{m}}'_i) + d^2(\tilde{\mathbf{m}}'_i, \mathbf{F}^T\tilde{\mathbf{m}}_i) \right) \ , \tag{15}$$

which operates simultaneously in the two images.

Let $\mathbf{l}_i = \mathbf{F}\tilde{\mathbf{m}}'_i \equiv [l_1, l_2, l_3]^T$ and $\mathbf{l}'_i = \mathbf{F}^T\tilde{\mathbf{m}}_i \equiv [l'_1, l'_2, l'_3]^T$. Using (14) and the fact that $\tilde{\mathbf{m}}_i^T\mathbf{F}\tilde{\mathbf{m}}'_i = \tilde{\mathbf{m}}_i'^T\mathbf{F}^T\tilde{\mathbf{m}}_i$, the criterion (15) can be rewritten as:

$$\min_{\mathbf{F}} \sum_i w_i^2(\tilde{\mathbf{m}}_i^T\mathbf{F}\tilde{\mathbf{m}}'_i)^2 \ , \tag{16}$$

where

$$w_i = \left(\frac{1}{l_1^2 + l_2^2} + \frac{1}{l_1'^2 + l_2'^2}\right)^{1/2} = \left(\frac{l_1^2 + l_2^2 + l_1'^2 + l_2'^2}{(l_1^2 + l_2^2)(l_1'^2 + l_2'^2)}\right)^{1/2} \ .$$

We now present two methods for solving this problem.

3.4.1 Iterative Linear Method

The similarity between (16) and (9) conducts us to solve the above problem by a *weighted* linear least-squares technique. Indeed, if we can compute the weight w_i for each point match, the corresponding linear equation can be multiplied by w_i (which is equivalent to replacing $\mathbf{u}_i$ in (8) by $w_i\mathbf{u}_i$), and exactly the same 8-point algorithm can be run to estimate the fundamental matrix, which minimizes (16).

The problem is that the weights w_i depends themselves on the fundamental matrix. To overcome this difficulty, we apply an iterative linear method. We first assume that all $w_i = 1$ and run the 8-point algorithm to obtain an initial estimation of the fundamental matrix. The weights w_i are then computed from this initial solution. The weighted linear least-squares is then run for an improved solution. This procedure can be repeated several times.

Although this algorithm is simple to implement and minimizes a physical quantity, our experience shows that there is no significant improvement compared to the original linear method. The main reason is that the rank-2 constraint of the fundamental matrix is not taken into account.

3.4.2 Nonlinear Minimization in Parameter Space

From the above discussions, it is clear that the right thing to do is to search for a matrix among the 3×3 matrices of rank 2 which minimizes (16). There are several possible parameterizations for the fundamental matrix (Luong 1992), e.g. we can express one row (or column) of the fundamental matrix as the linear combination of the other two rows (or columns). The parameterization described below is based directly on the parameters of the epipolar transformation (see Sect. 2.2).

Parameterization of fundamental matrix. Denoting the columns of $\mathbf{F}$ by the vectors $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$, we have:

$$\begin{array}{rcl} \mathrm{rank}(\mathbf{F}) & = & 2 \\ & \Longleftrightarrow & \\ (\exists j_0, j_1, j_2 \in [1,3])\ (\exists \lambda_1, \lambda_2 \in \mathcal{R}), & & \mathbf{c}_{j_0} + \lambda_1 \mathbf{c}_{j_1} + \lambda_2 \mathbf{c}_{j_2} = 0 \qquad (17) \\ (\nexists \lambda \in \mathcal{R}), & & \mathbf{c}_{j_1} + \lambda \mathbf{c}_{j_2} = 0\ . \qquad (18) \end{array}$$

Condition (18), as a non-existence condition, cannot be expressed by a parameterization: we shall only keep condition (17) and so extend the parameterized set to all the 3×3-matrices of rank strictly less than 3. Indeed, the rank-2 matrices of, for example, the following forms:

$$\left[\ \mathbf{c}_1 \quad \mathbf{c}_2 \quad \lambda \mathbf{c}_2\ \right] \text{ and } \left[\ \mathbf{c}_1 \quad \mathbf{0}_3 \quad \mathbf{c}_3\ \right] \text{ and } \left[\ \mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{0}_3\ \right]$$

do not have any parameterization if we take $j_0 = 1$. A parameterization of $\mathbf{F}$ is then given by $(\mathbf{c}_{j_1}, \mathbf{c}_{j_2}, \lambda_1, \lambda_2)$. This parameterization implies to divide the parameterized set among three maps, corresponding to $j_0 = 1$, $j_0 = 2$ and $j_0 = 3$.

If we construct a 3-vector such that λ_1 and λ_2 are the $j_1{}^{\text{th}}$ and $j_2{}^{\text{th}}$ coordinates and 1 is the $j_0{}^{\text{th}}$ coordinate, then it is obvious that this vector is the eigenvector of $\mathbf{F}$, and is thus the epipole in the case of the fundamental matrix. Using such a parameterization implies to compute directly the epipole which is often a useful quantity, instead of the matrix itself.

To make the problem symmetrical and since the epipole in the other image is also worth being computed, the same decomposition as for the columns is used for the rows, which now divides the parameterized set into 9 maps, corresponding to the choice of a column and a row as linear combinations of the two columns and two rows left. A parameterization of the matrix is then formed by the two coordinates x and y of the first epipole, the two coordinates x' and y' of the second epipole and the four elements a, b, c and d left by $\mathbf{c}_{i_1}$, $\mathbf{c}_{i_2}$, $\mathbf{l}_{j_1}$ and $\mathbf{l}_{j_2}$, which in turn parameterize the epipolar transformation mapping an epipolar line of the second image to its corresponding epipolar line in the first image. In that way, the matrix is written, for example, for $i_0 = 3$ and $j_0 = 3$:

$$\mathbf{F} = \begin{bmatrix} a & b & -ax' - by' \\ c & d & -cx' - dy' \\ -ax - cy & -bx - dy & (ax' + by')x + (cx' + dy')y \end{bmatrix}\ . \qquad (19)$$

At last, to take into account the fact that the fundamental matrix is defined only up to a scale factor, the matrix is normalized by dividing the four elements (a, b, c, d) by the largest in absolute value. We have thus in total 36 maps to parameterize the fundamental matrix.

Choosing the best map. Giving a matrix $\mathbf{F}$ and the epipoles, or an approximation to it, we must be able to choose, among the different maps of the parameterization, the most suitable for $\mathbf{F}$. Denoting by $\mathbf{f}_{i_0 j_0}$ the vector of the elements of $\mathbf{F}$ once decomposed as in equation (19), i_0 and j_0 are chosen in order to maximize the rank of the 9×8 Jacobian

matrix:

$$\mathbf{J} = \frac{d\mathbf{f}_{i_0 j_0}}{d\mathbf{p}} \quad \text{where} \quad \mathbf{p} = [x, y, x', y', a, b, c, d]^T \ . \tag{20}$$

This is done by maximizing the norm of the vector whose coordinates are the determinants of the nine 8×8 submatrices of $\mathbf{J}$. An easy calculation shows that this norm is equal to

$$(ad - bc)^2 \sqrt{x^2 + y^2 + 1} \sqrt{x'^2 + y'^2 + 1} \ .$$

At the expense of dealing with different maps, the above parameterization works equally well whether the epipoles are at infinity or not. This is not the case with the original proposition in (Luong 1992). More details can be found in (Csurka, Zeller, Zhang and Faugeras 1996).

Minimization. The minimization of (16) can now be performed by any minimization procedure. The Levenberg-Marquardt method (as implemented in `MINPACK` from `NETLIB` (More 1977) and in the Numeric Recipes in C (Press, Flannery, Teukolsky and Vetterling 1988)) is used in our program. During the process of minimization, the parameterization of $\mathbf{F}$ can change: The parameterization chosen for the matrix at the beginning of the process is not necessarily the most suitable for the final matrix. The nonlinear minimization method demands an initial estimate of the fundamental matrix, which is obtained by running the 8-point algorithm.

3.5 Gradient-based technique

Let $f_i = \widetilde{\mathbf{m}}_i^T \mathbf{F} \widetilde{\mathbf{m}}'_i$. Minimizing $\sum_i f_i^2$ does not yield a good estimation of the fundamental matrix, because the variance of each $\mathbf{f}_i$ is not the same. The least-squares technique produces an optimal solution if each term has the same variance. Therefore, we can minimize the following weighted sum of squares:

$$\min_{\mathbf{F}} \sum_i f_i^2 / \sigma_{f_i}^2 \ , \tag{21}$$

where $\sigma_{f_i}^2$ is the variance of f_i, and its computation will be given shortly. This criterion now has the desirable property: f_i / σ_{f_i} follows, under the first order approximation, the standard Gaussian distribution. In particular, all f_i / σ_{f_i} have the same variance, equal to 1. The same parameterization of the fundamental matrix as that described in the previous section is used.

Because points are extracted independently by the same algorithm, we make a reasonable assumption that the image points are corrupted by independent and identically distributed Gaussian noise, i.e. their covariance matrices are given by

$$\mathbf{\Lambda}_{\mathbf{m}_i} = \mathbf{\Lambda}_{\mathbf{m}'_i} = \sigma^2 \ \mathrm{diag}\,(1, 1) \ ,$$

where σ is the noise level, which may be not known. Under the first order approximation, the variance of f_i is then given by

$$\begin{aligned}\sigma_{f_i}^2 &= \frac{\partial f_i}{\partial \mathbf{m}_i}^T \mathbf{\Lambda}_{\mathbf{m}_i} \frac{\partial f_i}{\partial \mathbf{m}_i} + \frac{\partial f_i}{\partial \mathbf{m}_i'}^T \mathbf{\Lambda}_{\mathbf{m}_i'} \frac{\partial f_i}{\partial \mathbf{m}_i'} \\ &= \sigma^2[l_1^2 + l_2^2 + l_1'^2 + l_2'^2] \,,\end{aligned}$$

where $\mathrm{l}_i = \mathbf{F}\widetilde{\mathbf{m}}_i' \equiv [l_1, l_2, l_3]^T$ and $\mathrm{l}_i' = \mathbf{F}^T\widetilde{\mathbf{m}}_i \equiv [l_1', l_2', l_3']^T$. Since multiplying each term by a constant does not affect the minimization, the problem (21) becomes

$$\min_{\mathbf{F}} \sum_i (\widetilde{\mathbf{m}}_i^T \mathbf{F} \widetilde{\mathbf{m}}_i')^2 / g_i^2 \,,$$

where $g_i = \sqrt{l_1^2 + l_2^2 + l_1'^2 + l_2'^2}$ is simply the gradient of f_i. Note that g_i depends on $\mathbf{F}$.

It is shown (Luong 1992) that f_i/g_i is a first order approximation of the orthogonal distance from $(\mathbf{m}_i, \mathbf{m}_i')$ to the quadratic surface defined by $\widetilde{\mathbf{m}}^T \mathbf{F} \widetilde{\mathbf{m}}' = 0$.

3.6 Nonlinear Method Minimizing Distances Between Observation and Reprojection

If we can assume that the coordinates of the observed points are corrupted by additive noise and that the noises in different points are independent but with equal standard deviation (the same assumption as that used in the previous technique), then the maximum likelihood estimation of the fundamental matrix is obtained by minimizing the following criterion:

$$\mathcal{F}(\mathbf{f}, \mathtt{M}) = \sum_i (\|\mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i)\|^2 + \|\mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i)\|^2) \,, \tag{22}$$

where $\mathbf{f}$ represents the parameter vector of the fundamental matrix such as the one described in Sect. 3.4, $\mathtt{M} = [\mathtt{M}_1^T, \ldots, \mathtt{M}_n^T]^T$ are the structure parameters of the n points in space, while $\mathbf{h}(\mathbf{f}, \mathtt{M}_i)$ and $\mathbf{h}'(\mathbf{f}, \mathtt{M}_i)$ are the projection functions in the first and second image for a given space coordinates $\mathtt{M}_i$ and a given fundamental matrix between the two images represented by vector $\mathbf{f}$. Simply speaking, $\mathcal{F}(\mathbf{f}, \mathtt{M})$ is the sum of squared distances between observed points and the *reprojections* of the corresponding points in space. This implies that we estimate not only the fundamental matrix but also the structure parameters of the points in space. The estimation of the structure parameters, or *3D reconstruction*, in the uncalibrated case is an important subject and needs a separate section to describe it in sufficient details (see Sect. A). In the remaining subsection, we assume that there is a procedure available for 3D reconstruction.

A generalization to (22) is to take into account different uncertainties, if available, in the image points. If a point $\mathbf{m}_i$ is assumed to be corrupted by a Gaussian noise with mean zero and covariance matrix $\mathbf{\Lambda}_{\mathbf{m}_i}$ (a 2×2 symmetric positive-definite matrix), then the maximum likelihood estimation of the fundamental matrix is obtained by minimizing the following

criterion:

$$\mathcal{F}(\mathbf{f}, \mathtt{M}) = \sum_i \left(\Delta\mathbf{m}_i^T \mathbf{\Lambda}_{\mathbf{m}_i}^{-1} \Delta\mathbf{m}_i + \Delta\mathbf{m}_i'^T \mathbf{\Lambda}_{\mathbf{m}_i'}^{-1} \Delta\mathbf{m}_i' \right)$$

with

$$\Delta\mathbf{m}_i = \mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i) \quad \text{and} \quad \Delta\mathbf{m}_i' = \mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i) \ .$$

Here we still assume that the noises in different points are independent, which is quite reasonable.

When the number of points n is large, the nonlinear minimization of $\mathcal{F}(\mathbf{f}, \mathtt{M})$ should be carried out in a huge parameter space ($3n + 7$ dimensions because each space point has 3 degrees of freedom), and the computation is very expensive. As a matter of fact, we can separate the structure parameters from the fundamental matrix such that the optimization of the structure parameters is conducted in each optimization iteration for the parameters of the fundamental matrix, that is:

$$\min_{\mathbf{f}} \left\{ \sum_i \min_{\mathtt{M}_i} \left(\|\mathbf{m}_i - \mathbf{h}(\mathbf{f}, \mathtt{M}_i)\|^2 + \|\mathbf{m}_i' - \mathbf{h}'(\mathbf{f}, \mathtt{M}_i)\|^2 \right) \right\} \ . \tag{23}$$

Therefore, a problem of minimization over $(3n + 7)$-D space (22) becomes a problem of minimization over 7-D space, in the latter each iteration contains n independent optimizations of 3 structure parameters. The computation is thus considerably reduced. As will be seen in Sect. A, the optimization of structure parameters is nonlinear. In order to speed up still more the computation, it can be approximated by an analytic method; when this optimization procedure converges, we then restart it with the nonlinear optimization method.

The idea underlying this method is already well known in motion and structure from motion (Faugeras 1993, Zhang 1995) and camera calibration (Faugeras 1993). Similar techniques have also been reported for uncalibrated images (Mohr, Veillon and Quan 1993, Hartley 1993).

3.7 Robust Methods

Up to now, we assume that point matches are given. They can be obtained by techniques such as correlation and relaxation (Zhang, Deriche, Faugeras and Luong 1995). They all exploit some *heuristics* in one form or another, for example, intensity similarity or rigid/affine transformation in image plane, which are not applicable to most cases. Among the matches established, we may find two types of *outliers* due to

bad locations. In the estimation of the fundamental matrix, the location error of a point of interest is assumed to exhibit Gaussian behavior. This assumption is reasonable since the error in localization for most points of interest is small (within one or two

pixels), but a few points are possibly incorrectly localized (more than three pixels). The latter points will severely degrade the accuracy of the estimation.

false matches. In the establishment of correspondences, only heuristics have been used. Because the only geometric constraint, i.e., the epipolar constraint in terms of the *fundamental matrix*, is not yet available, many matches are possibly false. These will completely spoil the estimation process, and the final estimate of the fundamental matrix will be useless.

The outliers will severely affect the precision of the fundamental matrix if we directly apply the methods described above, which are all least-squares techniques.

Least-squares estimators assume that the noise corrupting the data is of zero mean, which yields an *unbiased* parameter estimate. If the noise variance is known, a *minimum-variance* parameter estimate can be obtained by choosing appropriate weights on the data. Furthermore, least-squares estimators implicitly assume that the entire set of data can be interpreted by *only one parameter vector* of a given model. Numerous studies have been conducted, which clearly show that least-squares estimators are vulnerable to the violation of these assumptions. Sometimes even when the data contains only one bad datum, least-squares estimates may be completely perturbed. During the last three decades, many robust techniques have been proposed, which are not very sensitive to departure from the assumptions on which they depend.

Recently, computer vision researchers have paid much attention to the robustness of vision algorithms because the data are unavoidably error prone (Haralick 1986, Zhuang, Wang and Zhang 1992). Many the so-called *robust regression* methods have been proposed that are not so easily affected by outliers (Huber 1981, Rousseeuw and Leroy 1987). The reader is referred to (Rousseeuw and Leroy 1987, Chap. 1) for a review of different robust methods. The two most popular robust methods are the *M-estimators* and the *least-median-of-squares* (LMedS) method, which will be presented below. Recent works on the application of robust techniques to motion segmentation include (Torr and Murray 1993, Odobez and Bouthemy 1994, Ayer, Schroeter and Bigün 1994), and those on the recovery of the epipolar geometry include (Olsen 1992, Shapiro and Brady 1995, Torr 1995)

3.7.1 M-Estimators

Let r_i be the *residual* of the i^{th} datum, i.e. the difference between the i^{th} observation and its fitted value. The standard least-squares method tries to minimize $\sum_i r_i^2$, which is unstable if there are outliers present in the data. Outlying data give an effect so strong in the minimization that the parameters thus estimated are distorted. The M-estimators try to reduce the effect of outliers by replacing the squared residuals r_i^2 by another function of the residuals, yielding

$$\min \sum_i \rho(r_i) \ , \tag{24}$$

where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen to be less increasing than square. Instead of solving directly this problem, we can implement it as an iterated reweighted least-squares one. Now let us see how.

Let $\mathbf{p} = [p_1, \ldots, p_p]^T$ be the parameter vector to be estimated. The M-estimator of $\mathbf{p}$ based on the function $\rho(r_i)$ is the vector $\mathbf{p}$ which is the solution of the following p equations:

$$\sum_i \psi(r_i) \frac{\partial r_i}{\partial p_j} = 0 \ , \quad \text{for } j = 1, \ldots, p, \tag{25}$$

where the derivative $\psi(x) = \mathrm{d}\rho(x)/\mathrm{d}x$ is called the *influence function*. If now we define a *weight function*

$$w(x) = \frac{\psi(x)}{x} \ , \tag{26}$$

then Equation (25) becomes

$$\sum_i w(r_i) r_i \frac{\partial r_i}{\partial p_j} = 0 \ , \quad \text{for } j = 1, \ldots, p. \tag{27}$$

This is exactly the system of equations that we obtain if we solve the following iterated reweighted least-squares problem

$$\min \sum_i w(r_i^{(k-1)}) r_i^2 \ , \tag{28}$$

where the superscript $^{(k)}$ indicates the iteration number. The weight $w(r_i^{(k-1)})$ should be recomputed after each iteration in order to be used in the next iteration.

The influence function $\psi(x)$ measures the influence of a datum on the value of the parameter estimate. For example, for the least-squares with $\rho(x) = x^2/2$, the influence function is $\psi(x) = x$, that is, the influence of a datum on the estimate increases linearly with the size of its error, which confirms the non-robustness of the least-squares estimate. When an estimator is robust, it may be inferred that the influence of any single observation (datum) is insufficient to yield any significant offset (Rey 1983). There are several constraints that a robust M-estimator should meet:

- The first is of course to have a bounded influence function.
- The second is naturally the requirement of the robust estimator to be unique. This implies that the objective function of parameter vector $\mathbf{p}$ to be minimized should have a unique minimum. This requires that *the individual ρ-function is* convex *in variable* $\mathbf{p}$. This is necessary because only requiring a ρ-function to have a unique minimum is not sufficient. This is the case with maxima when considering mixture distribution;

the sum of unimodal probability distributions is very often multimodal. The convexity constraint is equivalent to imposing that $\frac{\partial^2 \rho(.)}{\partial \mathbf{p}^2}$ is non-negative definite.

- The third one is a practical requirement. Whenever $\frac{\partial^2 \rho(.)}{\partial \mathbf{p}^2}$ is singular, the objective should have a gradient, i.e. $\frac{\partial \rho(.)}{\partial \mathbf{p}} \neq \mathbf{0}$. This avoids having to search through the complete parameter space.

There are a number of different M-estimators proposed in the literature. The reader is referred to (Zhang 1996b) for a comprehensive review.

It seems difficult to select a ρ-function for general use without being rather arbitrary. The result reported in Sect. 4 uses Tukey function:

$$r_i = \begin{cases} \frac{c^2}{6}\left(1 - \left[1 - \left(\frac{r_i}{c\sigma}\right)^2\right]^3\right) & \text{if } |r_i| \leq c\sigma \\ (c^2/6) & \text{otherwise,} \end{cases}$$

where σ is some estimated standard deviation of errors, and $c = 4.6851$ is the tuning constant. The corresponding weight function is

$$w_i = \begin{cases} [1 - (x/c)^2]^2 & \text{if } |r_i| \leq c\sigma \\ 0 & \text{otherwise.} \end{cases}$$

Another commonly used function is the following tri-weight one:

$$w_i = \begin{cases} 1 & |r_i| \leq \sigma \\ \sigma/|r_i| & \sigma < |r_i| \leq 3\sigma \\ 0 & 3\sigma < |r_i| \ . \end{cases}$$

In (Olsen 1992, Luong 1992), this weight function was used for the estimation of the epipolar geometry.

Inherent in the different M-estimators is the simultaneous estimation of σ, the standard deviation of the residual errors. If we can make a good estimate of the standard deviation of the errors of good data (inliers), then data whose error is larger than a certain number of standard deviations can be considered as outliers. Thus, the estimation of σ itself should be robust. The results of the M-estimators will depend on the method used to compute it. The *robust standard deviation* estimate is related to the median of the absolute values of the residuals, and is given by

$$\hat{\sigma} = 1.4826[1 + 5/(n-p)] \ \underset{i}{\text{median}} \, |r_i| \ . \tag{29}$$

The constant 1.4826 is a coefficient to achieve the same efficiency as a least-squares in the presence of only Gaussian noise (actually, the median of the absolute values of random

numbers sampled from the Gaussian normal distribution $N(0,1)$ is equal to $\Phi^{-1}(\frac{3}{4}) \approx 1/1.4826$); $5/(n-p)$ (where n is the size of the data set and p is the dimension of the parameter vector) is to compensate the effect of a small set of data. The reader is referred to (Rousseeuw and Leroy 1987, page 202) for the details of these magic numbers.

Our experience shows that M-estimators are robust to outliers due to bad localization. They are, however, not robust to false matches, because they depend heavily on the initial guess, which is usually obtained by least-squares. This leads us to use other more robust techniques.

3.7.2 Least Median of Squares (LMedS)

The LMedS method estimates the parameters by solving the nonlinear minimization problem:

$$\min \underset{i}{\text{median}}\ r_i^2 \ .$$

That is, the estimator must yield the smallest value for the median of squared residuals computed for the entire data set. It turns out that this method is very robust to false matches as well as outliers due to bad localization. Unlike the M-estimators, however, the LMedS problem cannot be reduced to a weighted least-squares problem. It is probably impossible to write down a straightforward formula for the LMedS estimator. It must be solved by a search in the space of possible estimates generated from the data. Since this space is too large, only a randomly chosen subset of data can be analyzed. The algorithm which we have implemented (the original version was described in (Zhang, Deriche, Luong and Faugeras 1994, Deriche, Zhang, Luong and Faugeras 1994, Zhang, Deriche, Faugeras and Luong 1995)) for robustly estimating the fundamental matrix follows the one structured in (Rousseeuw and Leroy 1987, Chap. 5), as outlined below.

Given n point correspondences: $\{(\mathbf{m}_i, \mathbf{m}'_i) | i = 1, \ldots, n\}$, we proceed the following steps:

1. A Monte Carlo type technique is used to draw m random subsamples of $p = 7$ different point correspondences (recall that 7 is the minimum number to determine the epipolar geometry).

2. For each subsample, indexed by J, we use the technique described in Sect. 3.1 to compute the fundamental matrix $\mathbf{F}_J$. We may have at most 3 solutions.

3. For each $\mathbf{F}_J$, we can determine the median of the squared residuals, denoted by M_J, with respect to the whole set of point correspondences, i.e.,

$$M_J = \underset{i=1,\cdots,n}{\text{median}} [d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}_J \widetilde{\mathbf{m}}'_i) + d^2(\widetilde{\mathbf{m}}'_i, \mathbf{F}_J^T \widetilde{\mathbf{m}}_i)] \ .$$

   Here, the distances between points and epipolar lines are used, but we can use other error measures.

4. Retain the estimate $\mathbf{F}_J$ for which M_J is minimal among all m M_J's.

The question now is: *How do we determine m ?* A subsample is "good" if it consists of p good correspondences. Assuming that the whole set of correspondences may contain up to a fraction ε of outliers, the probability that at least one of the m subsamples is good is given by

$$P = 1 - [1 - (1 - \varepsilon)^p]^m \ . \tag{30}$$

By requiring that P must be near 1, one can determine m for given values of p and ε:

$$m = \frac{\log(1 - P)}{\log[1 - (1 - \varepsilon)^p]} \ .$$

In our implementation, we assume $\varepsilon = 40\%$ and require $P = 0.99$, thus $m = 163$. Note that the algorithm can be speeded up considerably by means of parallel computing, because the processing for each subsample can be done independently.

As noted in (Rousseeuw and Leroy 1987), the LMedS *efficiency* is poor in the presence of Gaussian noise. The efficiency of a method is defined as the ratio between the lowest achievable variance for the estimated parameters and the actual variance provided by the given method. To compensate for this deficiency, we further carry out a weighted least-squares procedure. The *robust standard deviation* estimate is given by (29), that is,

$$\hat{\sigma} = 1.4826[1 + 5/(n - p)]\sqrt{M_J} \ ,$$

where M_J is the minimal median estimated by the LMedS. Based on $\hat{\sigma}$, we can assign a weight for each correspondence:

$$w_i = \begin{cases} 1 & \text{if } r_i^2 \leq (2.5\hat{\sigma})^2 \\ 0 & \text{otherwise} \ , \end{cases}$$

where

$$r_i^2 = d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}\widetilde{\mathbf{m}}'_i) + d^2(\widetilde{\mathbf{m}}_i, \mathbf{F}^T\widetilde{\mathbf{m}}'_i) \ .$$

The correspondences having $w_i = 0$ are outliers and should not be further taken into account. We thus conduct an additional step:

5. Refine the fundamental matrix $\mathbf{F}$ by solving the weighted least-squares problem:

$$\min \sum_i w_i r_i^2 \ .$$

The fundamental matrix is now robustly and accurately estimated because outliers have been detected and discarded by the LMedS method.